

# Non-destructive seed detection in mandarins: Comparison of automatic threshold methods in FLASH and COMSPIRA MRIs

P. Barreiro<sup>1</sup>, C. Zheng<sup>2</sup>, Da-Wen Sun<sup>3</sup>, N. Hernández-Sánchez<sup>1</sup>,  
J.M. Pérez-Sánchez<sup>1</sup>, J. Ruiz-Cabello<sup>1</sup>

*LPF-TAG, Departamento de Ingeniería Rural, ETSI Agrónomos, Universidad Politécnica de Madrid, Avda. Complutense s/n, 28.040 Madrid, Spain*  
*FRCFT, Biosystems Engineering, University College Dublin, National University of Ireland, Earlsfort Terrace, Dublin 2, Ireland*  
*Instituto de Estudios Biofuncionales, Universidad Complutense de Madrid, P<sup>o</sup> Juan XXIII, 1. 28.040 Madrid, Spain*

---

## Abstract

Magnetic resonance imaging was used to acquire images of the internal structure of mandarins for non-destructive seed identification. Two different types of fast MRI sequences were investigated: a gradient echo and a spiral–radial, with 484 ms acquisition time for the former compared to 240 ms for the latter. The radial–spiral option allows over-sampling of the central area of the  $k$ -space maintaining the contrast within the MRI images and so the feasibility of seed segmentation. Three segmentation techniques were applied for image post-processing: region-based, one-dimension histogram variance, and two-dimension histogram variance, among which the latter procedure has been demonstrated to give the most promising results. Image features including perimeter, compactness, maximum distance to the gravity centre, and aspect ratio were employed in a linear discriminant function, by which seed identification of mandarins could be achieved with 100% accuracy using radial–spiral sequence and 98.7% accuracy with gradient echo images.

**Keywords:** Image analysis; Classification; Citrus; Internal quality; Fruit

---

## 1. Introduction

There is a variety of image acquisition techniques available for the analysis of internal structures in food: ultrasound (amplitude modulation and brightness modulation), magnetic resonance imaging (MRI), computed tomography (based on X-ray attenuation), and electrical tomography (based on resistance, impedance, capacitance or electromagnetic inductance) (Du and Sun, 2004). Among these, MRI has the advantage of allowing a variety of measurements to be made that not only contribute to the evaluation of maturity and quality parameters in fruit and vegetables but also improve the understanding of underlying physiological processes (Butz et al., 2005). As stated by McCarthy (1994), typical experiments measure a component density, diffusivity and mobility, bulk flow of a fluid, or provide a measure of the internal structure of a system. For further details on this technique and its

applicability to food science see McCarthy (1994) and Hills (1998).

MRI has already been shown to be an effective technique for internal quality assessment in a wide variety of fruit species: apple, avocado, blueberry, cherimoya, courgette, cucumber, durian, kiwifruit, mandarin, mango, melon, nectarine, olive, onion, orange, papaya, pear, peach, pineapple, potato, tangerine, tomato, strawberry, melon, and watermelon (Hills and Clark, 2003). However, most of these investigations have been undertaken with commercial instruments designed for medical purposes, which operate at high magnetic field strength with reduced constraints on scan times. Time consuming processes (needed for acquisition and post-processing) are important issues for on-line quality evaluation in the food industry. Early MRI applications in the food industry have been performed to address this variable (Zion et al., 1994, 1997; Chen et al., 1996; Kim et al., 1999), although these studies focused on real-time Magnetic Resonance projections and not 2D imaging, which is limited by its longer acquisition times. A different approach, demonstrated for the evaluation of internal browning in apples, is provided by Chayaprasert and Stroschine

## Nomenclature

$C$	compactness
$i,j,k,l$	index parameters
$L$	maximum intensity
$N$	number of element in the set
$p$	probability distribution
$t$	threshold
$x,y$	coordinates
$x',y'$	centre of mass

## Greek letters

$\mu$	first moment
$\sigma$	variance
$\omega$	zeroth moment

## Subscripts

$A,B$	image classes
$i$	index parameter

(2005). These authors made use of relaxation signals from Carr–Purcell–Meiboom–Gill (CPMG) decay curves instead of imaging protocols, following a similar procedure to that anticipated by Barreiro et al. (2002) in static samples for mealiness detection.

There is an increasing interest in making 2D MRI transferable for the food industry. From our point of view, the on-line implementation of this technique demands the fulfilment of three requirements: the use of ultrafast sequences (below 1 s), the acquisition of images under motion conditions in order to reach conveyor velocities close to the current ones (up to 2 m/s), and the use of appropriate technology for low magnetic field strengths, keeping in mind sample polarisation, magnet and probe designs, field homogeneity, compatibility with commercial graders and budgets as pointed out by Hills (1998) and Hills and Clark (2003). For instance, the conventional sequence based on gradient echoes fast low angle shot (FLASH), which can capture MRI images within less than 500 ms, has been accomplished under motion conditions and applied for detection of freeze injury in oranges, seeds in mandarins, and internal breakdown in pears (Hernandez-Sanchez et al., 2004, 2005, 2006, 2007). As demonstrated in these studies, image artefacts from dynamic MRI can be removed with an appropriate slice selection and proper mathematical algorithms. Moreover, motion corrected images have been found statistically undistinguishable from static ones, making the image processing algorithms directly transferable from one condition to another.

A new MRI sequence called combined spiral and radial acquisition (COMSPIRA) has recently been proposed (Rodríguez et al., 2004) for polarised gas imaging of lung with significant shortening of acquisition times and enhanced contrast as compared to FLASH, being much less sensitizing to motion artefacts. In this study, the new sequence was used for internal fruit quality assessment and compared with

the conventional gradient echo sequence. The increasing feasibility of MRI for fruit quality assessment highlights automatic image segmentation and features extraction. The analysis of pros and cons for the different sequences regarding these post-processing methods becomes of increasing interest.

An automatic segmentation procedure is referred to as the unsupervised or supervised partitioning of the image into non-overlapping regions with different image characteristics, e.g., intensity and texture. For the unsupervised mode the class label of the output regions are not provided *a priori* whereas for the supervised mode those values are. The segmentation techniques available may be grouped into four different categories: histogram-, region-, edge- and classification-based (Du and Sun, 2004).

The histogram-based methods are preferable when there are only two objects in the image, such as a seed within a matrix of pulp, while the others are used for images containing multiple objects. Besides, histogram-based algorithms have the advantage of being intrinsically simple and requiring less computation time (Zheng et al., 2006a).

For histogram-based methods, an optimal threshold is computed by an objective function such as variance or entropy maximization to segment the histogram and thus the image. As stated by Unay and Gosselin (2006), local techniques calculate different thresholds for each pixel within a neighbourhood, whereas global ones calculate one single threshold for the whole image. Besides, adaptive techniques calculate a new threshold value for each image. These authors obtained better performance by applying global thresholding for the segmentation of “Jonagold” apples when compared to local techniques (Unay and Gosselin, 2006).

A recent study uses global adaptive thresholding methods based on one-dimension (1D) and two-dimension (2D) histograms to segment the images of beef joints (raw, cooked, and cooled) automatically from the background (Zheng et al., 2006a). These authors illustrate that the latter is more effective and that the development of an improved fast recurring algorithm overcomes the higher demand of computing resources so that the two-dimension histogram becomes feasible for real time applications. This methodology has been implemented in the present study for seed detection.

Region-based segmentation techniques group pixels according to homogeneity criteria. Although region-based algorithms are computationally more expensive than histogram-based techniques, they do not require *a priori* information so that their use is also of interest for automated image processing (Yang, 1994; Hernandez-Sanchez et al., 2004, 2006).

The objective of this study was to compare the performance of several automatic histogram- and region-based thresholding segmentation techniques for seed detection in mandarins using two different fast MRI sequences: FLASH and COMSPIRA. In addition, the available classification algorithms could be improved with the inclusion of some new image features in order to achieve, hopefully, 100% of identification accuracy, which is demanded by the fruit industry.

## 2. Materials and methods

Sample preparation and MRI image acquisition were carried out at the Polytechnic University of Madrid (UPM), while image processing and data analysis were conducted in a joint collaboration at the University College Dublin, National University of Ireland.

### 2.1. Materials

Mandarins (*Citrus reticulata* Blanco  $\times$  *Citrus sinensis* (L.) Osb.) var. Nadorcott were provided in March 2006 by a distributor of citrus in Valencia (Spain). The fruit, 78 in total, were carefully selected by experts so to achieve a wide range of variability in size and seeds and production fields. Samples were imaged at room temperature (22 °C).

### 2.2. MRI equipment

MRI experiments were performed immediately after harvest on a Bruker BIOSPEC 47/40 spectrometer (Ettlingen, Germany) operating at 200 MHz. The bore of the magnet was horizontal, 147 cm long and a 40 cm in diameter, which was reduced to 26 cm when the gradients stack in place. A birdcage coil with an inner diameter of 20 cm was used in acquisition-reception mode.

### 2.3. FLASH sequence

Fast low angle shot sequence (FLASH) combines a low flip angle radio frequency excitation pulse with the acquisition of a gradient recalled echo by reversal of the frequency-encoding gradient. The short flip angles (typically 5–10°) were such that the residual longitudinal magnetization permits repeated and rapid sampling, since it is not necessary to wait for its recovery by relaxation. A single  $k$ -space line is sampled for each RF pulse. The absence of a 180° refocusing pulse allows the echo train to be moved closer to the excitation pulse and thus shorten the effective echo time (TE). As a result of these arrangements, FLASH allows a faster data acquisition than the corresponding spin echo. The residual transverse magnetization can be destroyed by applying an additional short and pulsed field gradient (a spoiling pulse) just before the next excitation pulse (Parikh, 1992).

In this study, proton density weighted images were acquired to enhance differences between juicy flesh and solid-like seed tissue. In such images, seed tissue appears as a hypointense signal, while flesh tissue appears as bright pixels. The sequence parameters used for seed identification were selected as those used in previous studies: recovery time ( $T_R$ ) = 7.6 ms, echo time ( $T_E$ ) = 2.5 ms, and flip angle = 10° (Hernandez-Sanchez et al., 2006).

Mandarins were placed inside the magnet bore with their central axis (calyx-stem axis) along the  $z$ -direction to obtain axial images (perpendicular to  $z$ -axis) displaying the transverse section of the central axis within the middle of the field of view (FOV). With this orientation, seed-containing mandarins will

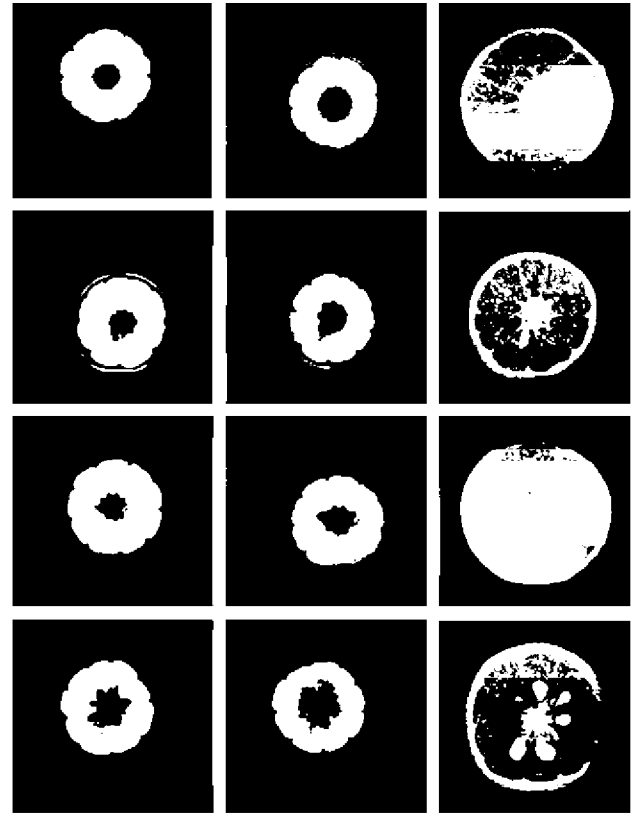


Fig. 1. Examples of FLASH and COMSPIRA MRI images (left and right, respectively) and corresponding RGB image on the cut section at approximately the centre of the FOV thickness. Acquisition details in the text.

have their central axis surrounded by a signal hypointense region corresponding to seed tissue. The FOV and the slice thickness used were 120 mm  $\times$  120 mm and 10 mm, respectively. The slice thickness was selected to acquire signals from the region with the highest probability of seed presence.

Images were collected with 64  $\times$  128 with a total acquisition time of 484 ms. All images were zero-filled to 128  $\times$  128 points before being reconstructed, and images were digitized at 4 bytes. Fig. 1 provides examples of FLASH images.

### 2.4. COMSPIRA sequence

Combined spiral radial (COMSPIRA) sequence samples the  $k$ -space from the origin with no gradient dephaser. After a slice selection (or non-slice-selective) radio frequency pulse, the procedure applies simultaneously a field gradient during which acquisition takes place. The radial projection-reconstruction version traverses  $k$  space in a radial way, with each radius being a straight line. In the spiral acquisition mode of COMSPIRA, single or multiple shot (or interleaved) archimedean trajectories are used dephased at a constant angle. In any of the two options, no echo procedure is required (Delpuech, 1995).

COMSPIRA makes an early acquisition of the low spatial frequencies in the  $k$ -space where the contrast information is contained, and thus it is possible to reduce the acquisition time without losing contrast between seeds and flesh.

The mathematical formulation behind COMSPIRA, allows one to choose between spiral and radial center-out  $k$ -space trajectories in two and three dimensions. For the acquisition in fruit, we have used a spiral version with the following parameters: 36  $k$ -space trajectories, the 512 points for trajectory, and 1 spirality factor (number of times the spiral revolves around the origin). The total acquisition time for these parameters is 240 ms.

The reduction from 484 ms for FLASH images to 240 ms when using COMSPIRA sequence would double the classification rate under on-line application.

Similar to the previous sequence, axial images were acquired with the spatial dimensions of the field of view being set to 120 mm  $\times$  120 mm and 10 mm of slice thickness. Images were digitized at 4 bytes and reconstructed to 128  $\times$  128 data points (see examples in Fig. 1). For COMSPIRA MRI, only 15 fruit were imaged among the 78 used for FLASH MRI due to limited NMR equipment availability. Those mandarins were selected by visual inspection of the FLASH images as representative of the global sample variability (5 mandarins with 0 seeds, 4 mandarins with 3 or less seeds, 6 mandarins with more than 3 seeds—including misclassified fruit by using FLASH images). Since more noise (computed as the average standard deviation of four background regions) is presented in the COMSPIRA images, a 3  $\times$  3 median filter was applied before further processing, which is expected not to alter the edges between regions.

## 2.5. Reference measurements

After MRI acquisition, mandarins were cut and RGB images were captured corresponding to the same slice as that of the axial views. The presence or absence of seeds was registered as the main discrimination criterion and target. Fig. 1 provides examples of non-destructive MRIs with corresponding RGB images of cut sections.

## 2.6. Image segmentation procedures

Three different segmentation techniques were evaluated in this study: the first one referred to as UPM (acronym of our University), was employed in our previous research (Hernandez-Sanchez et al., 2006), the second, dimension histogram variance thresholding (1DHVT) and the third, two dimension histogram variance thresholding (2DHVT) as proposed by colleagues at University College Dublin (Zheng et al., 2006a).

UPM algorithm is a region-based method that makes use of the 32 bit resolution data from MRI images of mandarins to automatically select a region of interest (ROI) by removing the background pixels. An initial threshold is set as the averaged maximum signal from the pixels in the ten first and last columns of the matrix image. This threshold is recomputed under an iterative process as the previous one multiplied by a variable factor. The iterative process is terminated whenever the difference of the ROI areas for two consecutive iterations is smaller than 2%, that is, stabilisation in the computed area is reached. Afterwards, the pixels within the hypointense region (HIR) are segmented within the ROI by a similar iterative procedure. The segmentation in this case uses only those pixels belonging to the ROI with larger iterative steps.

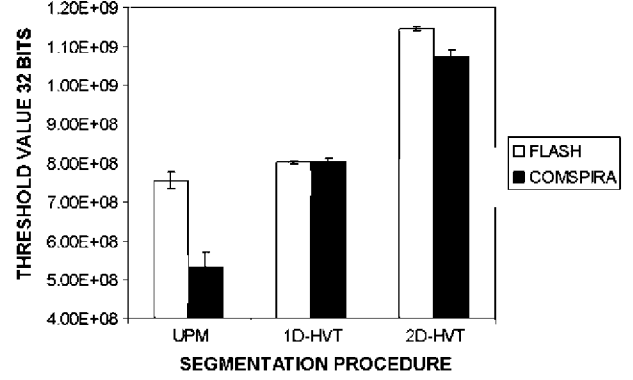


Fig. 2. Average threshold levels for the different MRI sequences: FLASH and COMSPIRA, as well as segmentation procedure: UPM, 1DHVT, and 2DHVT.

The other two segmentation techniques make use of 8 bit resolution images. Images therefore had to be scaled from 32 to 8 bits. For 1DHVT no a priori knowledge but the 8 bit intensity (gray level) histograms is needed (Fig. 2). The objective function employed is the maximization of the difference of intensity variance ( $\sigma$ ) between two resultant classes (A and B) as in the following equation:

$$\sigma(t)^2 = \omega_A(t)\omega_B(t)[\mu_B(t) - \mu_A(t)]^2 \quad (1)$$

where  $\omega$  and  $\mu$  are the zeroth and first cumulative moment of the histogram distribution which can be computed by the equations below:

$$\omega_A(t) = \sum_{i=1}^t p(i) \quad (2)$$

$$\omega_B(t) = \sum_{i=t+1}^L p(i) = 1 - \omega_A(t) \quad (3)$$

$$\mu_A(t) = \sum_{i=1}^t \frac{ip(i)}{\omega_A(t)} \quad (4)$$

$$\mu_B(t) = \sum_{i=t+1}^L \frac{ip(i)}{\omega_B(t)} \quad (5)$$

$p(i)$  being the normalised histogram and  $L$  the maximum number of gray levels (256).

In the 2DHVT a 2D histogram (Fig. 4) refers to the co-occurrence of two intensity values for the same pixel in two images: the original one and the averagely filtered image using a 3  $\times$  3 mask. Threshold segments the 2D histogram into four different classes: A, B, C, and D. Classes A and B are regarded as object and background or background and object, while classes C and D are known to contain the distributions of edges and noises, and thus, C and D are not considered (Zheng et al., 2006a). The objective function employed is the maximization of the intensity variance between the two resultant classes A and B computed using the following equations:

$$\sigma(k, l)^2 = \omega_A(k, l)\omega_B(k, l)[\mu_A(k, l) - \mu_B(k, l)]^2 \quad (6)$$

$$\omega_A(k, l) = \sum_{i=1}^k \sum_{j=1}^l p(i, j) \quad (7)$$

$$\omega_B(k, l) = \sum_{i=k+1}^L \sum_{j=l+1}^L p(i, j) \quad (8)$$

$$\mu_A(k, l) = \sum_{i=1}^k \sum_{j=1}^l \frac{ijp(i, j)}{\omega_A(k, l)} \quad (9)$$

$$\mu_B(k, l) = \sum_{i=k+1}^L \sum_{j=l+1}^L \frac{ijp(i, j)}{\omega_B(k, l)} \quad (10)$$

where  $p(i, j)$  is the normalised 2D histogram and  $L$  the maximum number of gray levels (256).

For the 2DHVT, the fast recurring algorithm developed by Zheng et al. (2006a) to accelerate the calculation of the zeroth and first moment of histogram was also employed.

### 2.7. Image features

As suggested by a previous study (Hernandez-Sanchez et al., 2006), perimeter ( $P$ ) and the maximum distance from the gravity centre to the boundary  $R_{\max}$  were used for the identification of seed-containing fruit. Apart from this, another three features, i.e., Second Order Moment selected so as to quantify the asymmetry of the hypointense-signal region corresponding to seeds and central axis (HIR), compactness of HIR (C), and aspect ratio (ASPECT) intended to address very elongated HIR, which would be typical for single seed fruit, were also proposed and calculated by the equations below:

$$C = 4\pi \text{HIR} \times P^{-2} \quad (11)$$

$$\text{ASPECT} = \frac{\text{MaxLength}}{\text{MaxWidth}} \quad (12)$$

$$\text{SOM} = \frac{1}{N} \sum_i (x_i - x')^2 (y_i - y')^2 \quad (13)$$

where  $(x_i, y_i)$  are the coordinates of the  $i$ th pixel in the HIR,  $(x', y')$  are coordinates of the centre of mass, and  $N$  is the number of pixels belonging to the HIR.

### 2.8. Data analysis

Analysis of variance has been conducted to test the statistical significance of the differences in the means of image features between segmentation methods and seedless versus seed-containing groups. This was accomplished by partitioning the total variance into the component, which is due to true random error, and the components, which are due to the differences between means. When no difference is found, the variance estimated based on within group variability should be about the same as the variance caused by between-groups variability. Those two estimates of variance can be compared via the Fisher test, which evaluates whether the ratio of the two variance estimates is significantly greater than 1 (StatSoft Inc, 2003).

As well, multivariate discriminant analysis was used to compute the classification functions, which were then used to determine the characteristics of the group, i.e., seed-containing or seedless, for each mandarin. For each group an individual classification function is developed and thus in total there are two classification functions. The function of each group computes the classification scores for each sample for the group, and after all the functions for a sample are computed, the sample is classified as belonging to a group for which the highest classification score is encountered (StatSoft Inc, 2003). In this study, previously available classification functions based on  $P$  and  $R_{\max}$  were used (Hernandez-Sanchez et al., 2006). Alternative functions were defined, which gathered new image features that are proposed within the manuscript. Finally, correlation analysis of image features was conducted so as to test their redundancy and complementary level.

## 3. Results and discussion

### 3.1. Comparison of MRI sequences

Fig. 1 provides examples of FLASH and COMSPIRA images from seedless, single seed, twin seed, and multiple seed mandarins. In all cases the presence of seeds is characterized by an increase in the hypointense region (HIR) together with some morphological distortions in such areas from the FLASH to COMSPIRA images. Both types of images show a good contrast between flesh and the HIR, which corresponds to central axis and seeds. It is interesting to note that for COMSPIRA images there is a significant loss in the boundary definition of the HIR even though the high contrast remains unchanged with regard to the FLASH MRIs. This fact is associated with the sequence procedure, as there is an early acquisition of the central area of the  $k$ -space where the contrast information is contained, while the outer  $k$ -space areas associated with spatial details are under-sampled. Furthermore, a higher level of noise was found in the background of COMSPIRA images (average STD of 4.86E7 a.u. compared to 8.0190E6 a.u. for FLASH images), which is also caused by the subsampling of low contrast areas in the  $k$ -space.

The higher level of noise found for COMSPIRA compared to FLASH images, probably was due to regridding methods used for reconstruction of these types of images, encouraged the use in the former of median filters prior to image segmentation, being useless for the latter. Other research also indicates the benefit of reducing the level of noise in MRIs. Lammertyn et al. (2003a) made use of a low-pass 2D adaptive noise-removal filter on MRIs before segmentation of core breakdown in pears, which was used in a time course evaluation of the defect (Lammertyn et al., 2003b). Instead, in this research we made use of a  $3 \times 3$  median filter in order to avoid deviation in the signal intensity level within the image before and after filtering.

Collewet et al. (2004) have indicated the existence of high and differential sensitivity of the MR acquisition protocols (proton density or  $T_2$  weighted) with regard to gray level normalisation methods, especially when texture features are to be obtained. Thybo et al. (2004) expressed the need for image normalisation when the gain level of the NMR scanner is unknown and

Table 1  
Fisher values for FLASH images from ANOVA to address the effect of the segmentation procedure and the influence of seeds on: intensity threshold (THR), hypointense region (HIR), perimeter ( $P$ ), maximum radius ( $R_{\max}$ ) second order moment (SOM), compactness ( $C$ ) and aspect ratio (ASPECT)

	THR	HIR	$P$	$R_{\max}$	SOM	$C$	ASP
Segmentation ( $N=78$ )	1376.3	4.9	3.6	5.3	5.2	ns	ns
Seed/no seed ( $N_S=24$ ; $N_{NS}=54$ )	ns	93.5	265.7	324.4	150.2	485.2	136.0
Interaction	ns	ns	ns	ns	ns	ns	ns

ns stands for no significant effect;  $N$  stands for the number of images;  $N_S$  stands for the number of images from seed-containing mandarins;  $N_{NS}$  stands for the number of images from seedless mandarins.

variable between work sessions, which was not the case in our study.

### 3.2. Comparison of segmentation procedures

Analysis of variance of the effect of segmentation procedure and presence of seeds has been carried out on the threshold value as well as on the image features computed from the FLASH (Table 1) and COMSPIRA images (data not shown). The Fisher value shows a substantial effect derived from the segmentation procedure applied ( $F=1376.5$  for FLASH and  $F=109.6$  for COMSPIRA), while no significant effect was found between seedless and seed-containing fruit.

Fig. 2 shows the average values of threshold expressed in 32 bit resolution for both FLASH and COMSPIRA. UPM threshold was always the lowest (averagely  $7.5E08$  and  $5.3E08$  for FLASH and COMSPIRA, respectively) while 2DHVT the highest ( $11.4E08$  and  $10.7E08$  for FLASH and COMSPIRA, respectively). UPM threshold is very sensitive to the type of sequence, while 1DHVT is not.

It is worth noting that there was a significant shift in the threshold value between 1DHVT and 2DHVT methods for both FLASH and COMSPIRA images. Figs. 3 and 4, respectively, show the thresholds obtained by 1DHVT ( $7.8E08$  a.u.) and 2DHVT ( $11.07E08$  a.u.) on the FLASH image of fruit number 8. Both methods 1DHVT and 2DHVT, maximize between-class

variance, although only the latter minimizes the amount of pixels labelled as edges or noise. This fact is thought to be the cause of the differences observed.

Extracted image features show a higher effect from seed presence than that for the segmentation procedure (Table 1), which indicates that all of the features are indeed useful to identify seed-containing fruit. The Fisher value of second order moments (SOM) for seed detection (150.2) gave a lower performance than  $R_{\max}$  and  $P$  (324.4 and 265.7, respectively). Perimeter ( $P$ ), maximum radius ( $R_{\max}$ ), and SOM show significant effects on the segmentation procedure. This fact is most relevant while taking into account the fact that previous research (Hernandez-Sanchez et al., 2006) made use of  $P$  and  $R_{\max}$  for seed identification.

Table 2 shows the average values of the image features for FLASH MRIs according to the presence of seeds and the segmentation methods. In all three segmentation methods,  $P$ ,  $R_{\max}$ , ASPECT, and SOM increase for seed-containing fruit compared with seedless mandarins while compactness ( $C$ ) is the lowest whenever fruit contain seeds or not. The difference between seed and seedless fruit was always the highest for 2DHVT compared with both 1DHVT and UPM. Moreover, the standard deviation of 2DHVT features was generally smaller compared with that of 1DHVT and UPM. It is important to state that  $P$  and  $R_{\max}$  for the 2DHVT show values comparable to those found in previous research (Hernandez-Sanchez et al., 2006). Thus, 2DHVT gave an average perimeter of 103 pixels for seed-containing mandarins while previous research gave 97.9 pixels, also, the average  $R_{\max}$  for seed-containing fruit under 2DHVT

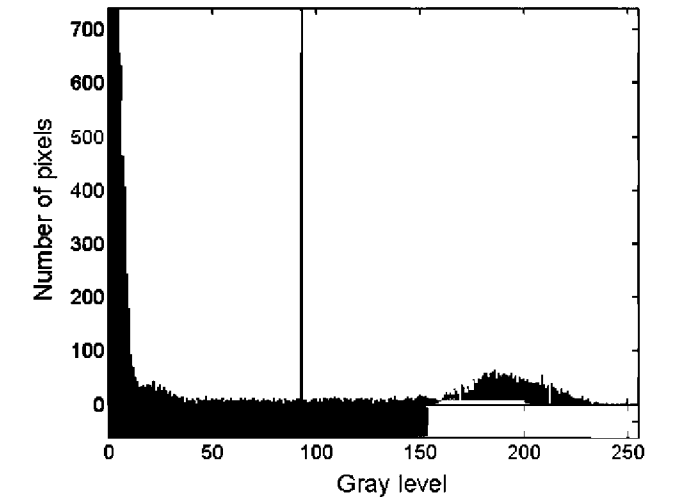


Fig. 3. Histogram of FLASH MRI of fruit number 8 and corresponding threshold level attained with 1DHVT procedure ( $92$  gray level under 8 bit resolution,  $7.8E08$  in 32 bit intensity levels).

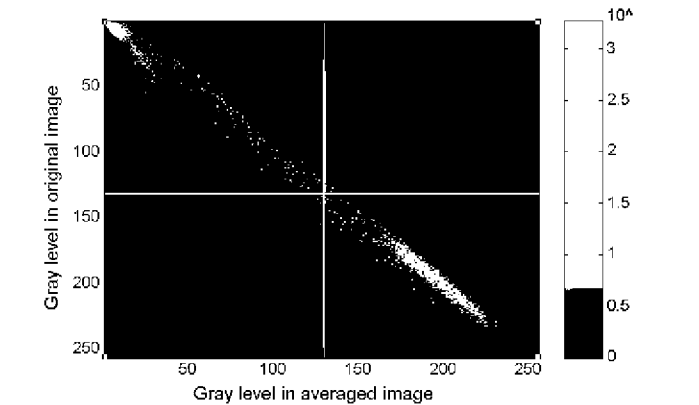


Fig. 4. Visualisation of co-occurrence matrix for FLASH image of fruit number 8 (colour scale from 0 to 1000). 2DHVT 8 bit threshold levels are 130 and 132 for the original and average image, respectively ( $11.1E09$  and  $10.9E09$  in 32 bit intensity levels). Background class up-left and object class bottom-right; edges and noises classes up-right and bottom-left.

Table 2

Average and standard deviation for some features: perimeter ( $P$ ), maximum radius ( $R_{\max}$ ), compactness ( $C$ ) and aspect ratio (ASPECT)

	$P$		$R_{\max}$		$C$		ASPECT	
	$S$	NS	$S$	NS	$S$	NS	$S$	NS
UPM	<b>86.29</b> <i>28.41</i>	<b>52.09</b> <i>14.93</i>	<b>15.55</b> <i>3.73</i>	<b>10.30</b> <i>2.51</i>	<b>0.69</b> <i>0.21</i>	<b>1.11</b> <i>0.17</i>	<b>1.26</b> <i>0.19</i>	<b>1.12</b> <i>0.06</i>
1DHVT	<b>92.67</b> <i>30.99</i>	<b>53.63</b> <i>14.94</i>	<b>15.90</b> <i>3.29</i>	<b>10.43</b> <i>2.24</i>	<b>0.63</b> <i>0.20</i>	<b>1.11</b> <i>0.19</i>	<b>1.25</b> <i>0.20</i>	<b>1.11</b> <i>0.06</i>
2DHVT	<b>103.00</b> <i>31.40</i>	<b>57.54</b> <i>12.48</i>	<b>17.85</b> <i>3.21</i>	<b>11.39</b> <i>2.33</i>	<b>0.64</b> <i>0.18</i>	<b>1.13</b> <i>0.13</i>	<b>1.26</b> <i>0.17</i>	<b>1.09</b> <i>0.05</i>
$N$	24	54	24	54	24	54	24	54

$N$  stands for the number of FLASH images according to the presence of seeds;  $S$  for seed-containing and NS for seedless mandarins. Type of segmentation procedures: region-based (UPM), one dimension histogram variance thresholding (1DHVT) and two dimension histogram variance thresholding (2DHVT). Bold values denote the average and italic values denote the standard deviation.

reached 17.85 pixels compared to 17.7 from the previous study.

Table 3 shows the average values for  $R_{\max}$  and ASPECT in COMSPIRA MRIs according to the presence of seeds and the segmentation methods. As in Table 2, both parameters show higher values for seed-containing than for seedless fruit. 2DHVT gives the most similar results when comparing FLASH and COMSPIRA.

Lammertyn et al. (2003a) made use of gray level thresholding for core breakdown identification in ‘Conference’ pears, although a fixed value was used for all the samples after visual inspection of the images, instead of automated thresholding as discussed in the present paper.

### 3.3. Validation of available discriminant function

Previous research (Hernandez-Sanchez et al., 2006) had proposed a linear discriminant function based on image features extracted by applying the UPM method for image segmentation,  $P$  and  $R_{\max}$ , in order to identify seed-containing mandarins on FLASH images, allowing 92.5% identification of seedless fruit and 79.5% of seed-containing mandarins for 33 fruit in total (18

seedless and 15 seed-containing mandarins). When this function was directly applied to features extracted with the three segmentation methods on FLASH images (Table 4), the best results were found for 2DHVT on FLASH MRI with 94.5% of classified seedless fruit and 73.9% of seed-containing mandarins. Therefore, it can be concluded that the 2DHVT gives the most robust performance against external validation.

These results are not yet good enough for the citrus industry since one of the major concerns is the segregation of seed-containing fruit, mainly mandarins. Moreover, the challenge involves the identification of fruit with a single seed. Therefore, taking into account the fact that 2DHVT provides a hypointense area segmentation that matches the region of seeds and central axis identified by visual inspection, it was decided to evaluate the rest of the image features (SOM, ASPECT and  $C$ ) to improve the quality of the discriminant functions in FLASH images with the expectation of extending them to more noisy COMSPIRA images.

### 3.4. Correlation between image features

Table 5 shows the correlation matrix between the image features extracted from the FLASH images with 2DHVT (78 samples in total). As expected,  $P$  and  $R_{\max}$  were strongly correlated ( $r=0.93$ ), which is similar to previous research (Hernandez-Sanchez et al., 2006). The variables which are less

Table 3

Average and standard deviation for maximum radius ( $R_{\max}$ ) and aspect ratio (ASPECT) in COMSPIRA MRIs

	$R_{\max}$		ASPECT	
	$S$	NS	$S$	NS
UPM	<b>14.17</b> <i>2.50</i>	<b>12.10</b> <i>1.12</i>	<b>1.20</b> <i>0.07</i>	<b>1.06</b> <i>0.03</i>
1DHVT	<b>15.64</b> <i>2.29</i>	<b>12.57</b> <i>1.31</i>	<b>1.23</b> <i>0.11</i>	<b>1.05</b> <i>0.04</i>
2DHVT	<b>17.25</b> <i>2.00</i>	<b>13.88</b> <i>1.97</i>	<b>1.24</b> <i>0.10</i>	<b>1.06</b> <i>0.02</i>
$N$	10	5	10	5

$N$  stands for the number of images according to the presence of seeds;  $S$  for seed-containing and NS for seedless mandarins. Type of segmentation procedures: region-based (UPM), one dimension histogram variance thresholding (1DHVT) and two dimension histogram variance thresholding (2DHVT). Bold values denote the average and italic values denote the standard deviation.

Table 4

Percentage of well classified fruit obtained under external validation using a discriminant function based on perimeter ( $P$ ) and maximum radius ( $R_{\max}$ ) for FLASH images developed in a previous research study

	FLASH			
	UPM	1DHVT	2DHVT	$N$
Seed	45.8%	58.3%	73.9%	24
No seed	100.0%	100.0%	94.5%	54
Total	83.3%	87.2%	91.0%	78

Original data (18 and 15 FLASH images from seedless and seed-containing mandarins, respectively) provided 87.9% of well classified mandarins (79.5% seed-containing and 92.5% seedless) (Hernandez-Sanchez et al., 2006).  $N$  stands for the number of samples.

Table 5

Correlation matrix for image features: perimeter ( $P$ ), maximum radius ( $R_{\max}$ ) second order moment (SOM), compactness ( $C$ ) and aspect ratio (ASPECT) from FLASH images segmented with two dimension histogram variance thresholding (2DHVT)

	$P$	$C$	$R_{\max}$	SOM	ASPECT
$P$	1.00				
$C$	-0.82	1.00			
$R_{\max}$	0.93	-0.76	1.00		
SOM	0.95	-0.66	0.94	1.00	
ASPECT	0.25	-0.49	0.31	0.18	1.00

related to  $P$  and  $R_{\max}$ , are the aspect ratio (ASPECT,  $r=0.25$  and  $0.31$ ) and the compactness ( $C$ ,  $r=-0.49$  and  $-0.76$ ). This suggests that they should contain complementary information and thus help to improve the classification capability. On the contrary, SOM is very much redundant with regard to  $P$  and  $R_{\max}$ . That is to say no further contribution of this feature for seed classification is expected.

Very little information is found in the literature concerning automated feature extraction in MRI fruit applications, generally related with texture analysis (Letal et al., 2003; Thybo et al., 2004). Although image texture is another important feature (Zheng et al., 2006b), it is mostly relevant for the evaluation of the general appearance and spatial arrangement of features in an image. However, for cases such as seed identification where a particular object is to be addressed, texture is not as important as size or shape and thus not used here.

### 3.5. Optimisation of discriminant functions

New discriminant functions were built based on the features extracted from 2DHVT segmented FLASH and COMSPIRA images. The most efficient function having a high percentage of well classified fruit together with the lowest number of variables makes use of  $P$ ,  $R_{\max}$ , and ASPECT (Table 6). The classification functions developed for FLASH images are statistically indistinguishable from that for COMSPIRA. For FLASH 98.7% of fruit (78 in total) were correctly classified

Table 6

Percentage of well classified fruit when developing classification functions based on perimeter ( $P$ ), maximum radius ( $R_{\max}$ ) and aspect ratio (ASPECT)

	FLASH		COMSPIRA	
	2DHVT	$N$	2DHVT	$N$
Seed	95.8%	23	100.0%	10
No seed	100.0%	54	100.0%	5
Total	98.7%	78	100.0%	15

Both MRI sequences, FLASH and COMSPIRA, make use of the same classification functions.  $N$  stands for the number of samples.

while 100% of classification were found for COMSPIRA (15 in total).

Fig. 5 shows three dimension plots, consisting of  $P$ ,  $R_{\max}$ , and ASPECT, for the FLASH and COMSPIRA-acquired images. Compared with seed-containing fruit (squares), seedless mandarins (circles) corresponded to low  $P$  values together with low  $R_{\max}$  and low ASPECT. It is worth noting that single seed fruit have special features like boundary  $R_{\max}$  values (between 15 and 16 pixels) together with a high ASPECT ratio (above 1.3, i.e., very elongated HIR is found).

Using  $P$ ,  $R_{\max}$ ,  $C$ , and ASPECT (four variables instead of three), the same percentage of well classified fruit as for three variables was found: 98.7% for FLASH and 100% for COMSPIRA (Table 6). However, the mis-classified fruit from FLASH was different from the discriminant function that makes use of  $P$ ,  $R_{\max}$ , and ASPECT. That is to say, looking at the compactness as secondary decision support enhances the classification of certain types of individuals while worsening others. To achieve the desired 100% identification of seed-containing fruit in FLASH MRIs, it is possible to combine both discriminant functions together with simple fusion criteria: a fruit needs to be labelled seedless by both functions, while requiring only one label to be classified as a seed-containing fruit.

Even when using three image features, the error rates (1.3% for FLASH and 0% for COMSPIRA) may lie within the citrus industrial interest. There is a need for interaction between industry and researchers on this topic that would allow estab-

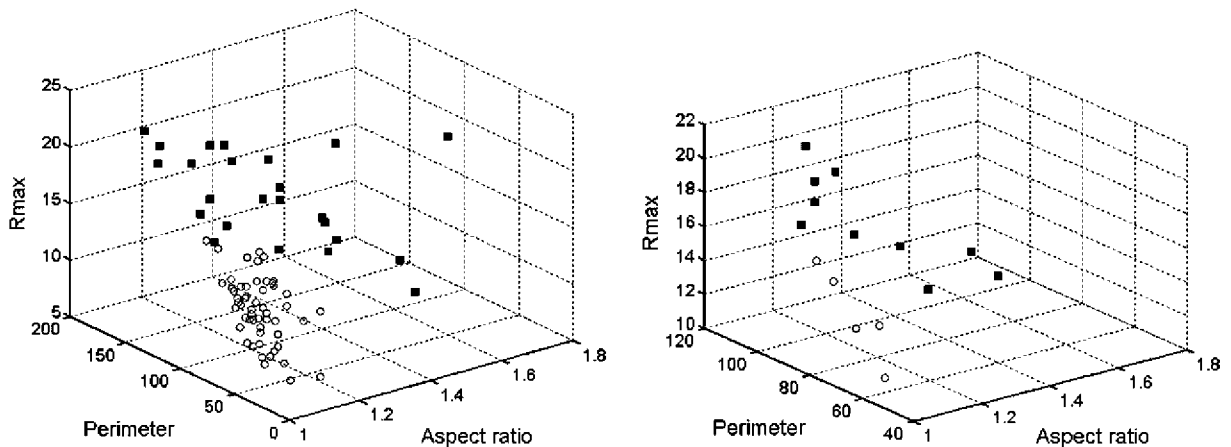


Fig. 5. Three dimension representation of mathematical features of FLASH images ( $N=78$ ) on the left and COMSPIRA images ( $N=15$ ) on the right. Squares (■) refer to seed-containing fruits while circles (○) indicate seedless mandarins.



lishing the targets susceptible of being accomplished since the classification performance in potential on-line NMR applications in the literature is very wide, ranging from less than 5% for pit detection in cherries and olives (Zion et al., 1994, 1997; Kim et al., 1999) to over 20% in the identification of internal browning in apples at high speed (Chayaprasert and Stroshine, 2005).

Classification performance in this study is very encouraging, taking into account that static FLASH images are indistinguishable from on-line MRI when the appropriate image plane with regard to motion is selected (Hernandez-Sanchez et al., 2005, 2006). However, the effect of sample movement on COMSPIRA images remains to be addressed. The use of oriented fruit in either FLASH or COMSPIRA images is basic in the performance of this application, which agrees with other studies (Kim et al., 1999; Chayaprasert and Stroshine, 2005). In future work, down shifting from 4.7 T towards low magnetic field strengths (lower than 1 T) would become a major issue in this application with the reduction in signal to noise ratio as the main hurdle to be faced when validating the performance of the segmentation algorithms.

#### 4. Conclusions

Two dimension histogram variance threshold (2DHVT) gives the best performance for seed segmentation compared to 1DHVT and other region-based methods. This procedure provides similar results in FLASH and COMSPIRA images. For the latter, background noise is magnified and thus the use of median filters is recommended before segmentation.

A previous available classification function has been validated with fruit from a new season achieving 91.0% of well classified fruit for 2DHVT FLASH images which is comparable to 87.9% of original data. However, only 73.9% of seed-containing fruit have been identified (compared to 79.5% for the original data). Results are still far from having industrial interest.

The improvement in image segmentation favours the definition of new morphological features for seed identification allowing 98.7% of well classified fruit for FLASH and 100% for COMSPIRA, respectively, 95.8% and 100% of seed-containing mandarins. These values become of much greater industrial interest.

One hundred percent of well classified seed-containing fruit can be obtained with both sequences when combining two discriminant functions and simple decision criteria. Such double check strategy will be externally validated in the near future.

#### Acknowledgements

This research was supported by the Spanish Ministry of Education in the framework of OPTISCAN (MCYT). It has been carried out under the scope of TAGRALIA-CAM supported by the Community of Madrid (Spain).

The authors would like to acknowledge MARTINAVARRO for the dedicated selection of the samples.

#### References

- Barreiro, P., Moya, A., Correa, E., Ruiz-Altisent, M., Fernández-Valle, M., Peirs, A., Wright, K.M., Hills, B.P., 2002. Prospect for the rapid detection of meakness in apples by nondestructive NMR relaxometry. *Appl. Magn. Reson.* 22, 387–400.
- Butz, P., Hofmann, C., Tauscher, B., 2005. Recent developments in noninvasive techniques for fresh fruit and vegetable internal quality analysis. *J. Food Sci.* 70, 131–141.
- Collewet, G., Strzelecki, M., Mariette, F., 2004. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn. Reson. Imaging* 22, 81–91.
- Chayaprasert, W., Stroshine, R., 2005. Rapid sensing of internal browning in whole apples using a low-cost, low-field proton magnetic resonance sensor. *Postharvest Biol. Technol.* 36, 291–301.
- Chen, P., McCarthy, M.J., Kim, S.M., Zion, B., 1996. Development of a high-speed NMR technique for sensing maturity of avocados. *Trans. ASAE* 39, 2205–2209.
- Du, C., Sun, D.-W., 2004. Recent developments in the applications of image processing techniques for food quality evaluation. *Trends Food Sci. Technol.* 15, 230–249.
- Delpuech, J.-J., 1995. *Dynamics of Solutions and Fluids Mixtures by NMR*. John Wiley & Sons, Chichester, U.K.
- Hernandez-Sanchez, N., Barreiro, P., Ruiz-Altisent, M., Ruiz-Cabello, J., Encarnación Fernández-Valle, M.E., 2004. Detection of freeze injury in oranges by magnetic resonance imaging of moving samples. *Appl. Magn. Reson.* 26, 431–445.
- Hernandez-Sanchez, N., Barreiro, P., Ruiz-Altisent, M., Ruiz-Cabello, J., Encarnación Fernández-Valle, M.E., 2005. Detection of seeds in citrus using MRI under motion conditions and improvement with motion correction. *Concepts Magn. Reson. Part B* 26, 81–92.
- Hernandez-Sanchez, N., Barreiro, P., Ruiz-Cabello, J., 2006. Detection of seeds in mandarins using MRI under motion conditions. *Biosyst. Eng.* 95, 529–536.
- Hernandez-Sanchez, N., Hills, B.P., Barreiro, P., Marigheto, N., 2007. An NMR study on internal browning in pears. *Postharvest Biol. Technol.* 44, 260–270.
- Hills, B., 1998. *Magnetic Resonance Imaging in Food Science*. John Wiley & Sons, New York.
- Hills, B.P., Clark, C.J., 2003. Quality assessment of horticultural products by NMR. *Ann. Rep. NMR Spectrosc.* 50, 75–120.
- Kim, S.M., Chen, P., McCarthy, M.J., Zion, B., 1999. Fruit internal quality evaluation using on-line nuclear magnetic resonance sensors. *J. Agric. Eng. Res.* 74, 293–301.
- Lammertyn, J., Dresselaers, T., Van Hecke, P., Jancsó, P., Wevers, M., Nicolai, B.M., 2003a. MRI and X-ray CT study of spatial distribution of core breakdown in ‘Conference’ pears. *Magn. Reson. Imaging* 21, 805–815.
- Lammertyn, J., Dresselaers, T., Van Hecke, P., Jancsó, P., Wevers, M., Nicolai, B.M., 2003b. Analysis of the time course of core breakdown in ‘Conference’ pears by means of MRI and X-ray CT. *Postharvest Biol. Technol.* 29, 19–28.
- Letal, J., Jirá, D., Šuderlová, L., Hájek, M., 2003. MRI ‘texture’ analysis of MR images of apples during ripening and storage. *LWT-Food Sci. Technol.* 36, 719–727.
- McCarthy, M.J., 1994. *Magnetic Resonance Imaging in Foods*. Chapman and Hall, New York.
- Parikh, A.M., 1992. *Magnetic Resonance Imaging Techniques*. Elsevier, New York, USA.
- Rodríguez, I., Pérez de Alejo, R., Cortijo, M., Ruiz-Cabello, J., 2004. COMSPIRA: a common approach to spiral and radial MRI. *Concepts Magn. Reson. Part B* 20, 40–44.
2003. STATISTICA (data analysis software system), Version 6. StatSoft, Inc., <http://www.statsoft.com>.
- Thybo, A.K., Szczypinski, P.M., Karlsson, A.H., Dønstrup, S., Stødkilde-Jørgensen, H.S., Andersen, H.J., 2004. Prediction of sensory texture quality attributes of cooked potatoes by NMR-imaging (MRI) of raw potatoes in combination with different image analysis methods. *J. Food Eng.* 61, 91–100.
- Unay, D., Gosselin, B., 2006. Automatic defect segmentation of ‘Jonagold’ apples on multi-spectral images: a comparative study. *Postharvest Biol. Technol.* 42, 271–279.

- Yang, Q., 1994. An approach to apple surface feature detection by machine vision. *Comput. Electron. Agric.* 11, 249–264.
- Zheng, C., Sun, D.-W., Zheng, L., 2006a. Segmentation of beef joint images using histogram thresholding. *J. Food Process Eng.* 29, 574–591.
- Zheng, C., Sun, D.-W., Zheng, L., 2006b. Recent applications of image texture for the evaluation of food qualities—a review. *Trends Food Sci. Technol.* 17, 113–128.
- Zion, B., McCarthy, M.J., Chen, P., 1994. Real-time detection of pits in processed cherries by magnetic resonance projections. *Lebensm. -Wiss. U. -Technol.* 27, 457–462.
- Zion, B., Kim, S.M., McCarthy, M.J., Chen, P.J., 1997. Detection of pits in olives under motion by nuclear magnetic resonance. *J. Sci. Food Agric.* 75, 496–502.